

Klasifikasi Tanaman Zaitun Menggunakan Algoritma K-Nearest Neighbor Dan Metode Gray Level Co-Occurrence Matrix

Ghenas Daffa S¹, Jayanta*², Ati Zaidiah³,
Program Studi Informatika / Fakultas Ilmu Komputer
Universitas Pembangunan Nasional “Veteran” Jakarta
Jl. R.S Fatmawati No. 1, Jakarta Selatan 12450
ghenasdaffa@protonmail.com

Abstrak. Zaitun merupakan tanaman yang berasal dari daerah Timur Tengah. Tanaman ini dapat hidup secara baik di Indonesia dikarenakan iklim Indonesia yang lumayan memiliki kesamaan dengan tempat asal tanaman ini yaitu sama-sama berupa iklim panas. Maka dari itu, sebetulnya Indonesia dapat menjadi tempat budidaya tanaman zaitun namun banyak masyarakat yang belum mengetahui bagaimana mengidentifikasi tanaman ini dengan tanaman lain. Oleh karena itu, penulis melakukan penelitian untuk membuat suatu program untuk mempermudah masyarakat dalam mengidentifikasi tanaman zaitun dengan tanaman lain melalui medium daun. Program ini akan menggunakan algoritma *K-Nearest Neighbor* (KNN) untuk model klasifikasi daun tanaman dan metode *Gray Level Co-occurrence Matrix* (GLCM) untuk ekstraksi ciri tekstur daun. Diharapkan penelitian ini dapat membantu untuk membedakan tanaman zaitun dengan tanaman lain.

Kata Kunci: daun, zaitun, *Grey Level Co-occurrence Matrix*, *K-Nearest Neighbor*

1 Pendahuluan

Zaitun adalah suatu tanaman buah yang banyak tumbuh di daerah beriklim panas. Memiliki nama latin *Olea europaea*, zaitun pada awalnya berasal dari daerah Timur Tengah dan sebagian Afrika [1]. Tanaman ini seringkali diolah buahnya yang sudah tua menjadi ekstrak yang mana kemudian dijadikan minyak. Di Indonesia, zaitun seringkali di pelihara sebagai tanaman hias. Sebagai tanaman yang tumbuh pada daerah panas, zaitun sangat subur ketika musim panas tiba. Zaitun juga merupakan tanaman yang sangat sensitif terhadap perubahan tempat dan bisa mati apabila akarnya masih belum cukup kuat. Maka dari itu tanaman ini sebetulnya tergolong sulit untuk dibudidayakan. Kendati demikian, zaitun memiliki banyak manfaat terutama buahnya. Selain dapat menjadi tanaman hias, buahnya dapat dijadikan bahan makanan atau diperas untuk diambil ekstraknya untuk dijadikan minyak zaitun. Namun, harga yang relatif mahal seringkali menyebabkan masyarakat enggan untuk membudidayakannya. Padahal tanaman ini juga dapat ditemukan di beberapa hutan secara liar namun masyarakat kurang memiliki informasi untuk membedakan jenis tanaman ini dengan tanaman lain sehingga tanaman ini kurang populer di masyarakat. Dengan adanya permasalahan tersebut, maka penulis memutuskan untuk melakukan penelitian berupa klasifikasi tanaman zaitun berdasarkan tekstur daun untuk membantu masyarakat dalam membedakan tanaman zaitun dengan tanaman lain yang mana dalam penelitian ini menggunakan tanaman delima.

Klasifikasi seringkali dilakukan secara manual yaitu melihat langsung objek yang bersangkutan. Pada praktiknya, hal ini seringkali memakan waktu dan kurang efektif ketika menghadapi objek berjumlah banyak. Belum lagi dibutuhkan sumber daya yang tidak sedikit. Namun pada masa sekarang, dengan berkembangnya teknologi, klasifikasi dapat dilakukan secara cepat dengan melakukan analisis ekstrak citra objek berdasarkan beberapa fitur. Fitur-fitur tersebut dapat berupa warna, tekstur, dan bentuk.

Adapun metode yang digunakan sebagai metode klasifikasi daun tanaman zaitun adalah algoritma K-Nearest Neighbor (KNN) sebagai metode klasifikasi citra dan Gray Level Co-occurrence Matrix sebagai metode ekstraksi fitur dari citra objek. Penelitian ini dilakukan dengan mengekstrak fitur tekstur pada daun yang dijadikan objek penelitian menggunakan GLCM, lalu menggunakan KNN sebagai algoritma klasifikasi fitur citra tersebut.

2 Persiapan Naskah

2.1 Tanaman Zaitun

Zaitun (*Olea europaea*) merupakan pohon kecil atau perdu yang berjenis tanaman hijau abadi atau *evergreen* dengan ketinggian antara 6-15 meter. Daun zaitun sendiri tersusun secara tunggal berhadapan tanpa adanya daun penumpu yang memiliki dimensi sekitar 20-90mm x 7-15mm dan memiliki bentuk daun lanset, tepi rata, permukaan atas yang licin, dan berwarna hijau keabu-abuan pada bagian atas serta warna kuning keemasan pada bagian bawah. Buahnya sendiri dapat dimakan mentah ketika masih muda, dan dapat diperas dan diekstrak menjadi minyak ketika sudah tua [2].

Tanaman zaitun memiliki buah kecil yang berupa buah batu, memiliki panjang sekitar 1–2,5 cm dengan biji yang memiliki endosperma. Pada zaitun liar, buahnya cenderung lebih kurus dan kecil dibanding zaitun budidaya. Zaitun mulai menghasilkan buah pada umur lima tahun dan usianya dapat mencapai ribuan tahun, yang mana dapat berukuran menjadi pohon besar. Buahnya dapat dipanen ketika masih hijau hingga sudah berwarna ungu tua.

2.2 Deteksi Tepi Canny

Canny merupakan sebuah algoritma deteksi tepi yang digunakan dalam analisis citra dalam bentuk ekstraksi ciri yang berfungsi untuk mendapatkan informasi berdasarkan ciri yang ditentukan dan mengambil fitur berdasarkan bentuk yang terdeteksi dalam tepi sebuah citra. Canny merupakan deteksi tepi optimal yang menggunakan *Gaussian Derivative Kernel* untuk menghilangkan derau dari citra agar mendapatkan hasil deteksi tepi yang halus [3].

2.3 Metode Grey Level Co-occurrence Matrix

Metode GLCM merupakan metode ekstraksi ciri statistik orde kedua dengan memakai matriks kookorens yang mempresentasikan hubungan ketetanggaan antara dua piksel pada orientasi sudut dan jarak tertentu. Sudut yang digunakan umumnya adalah sudut 0°, 45°, 90°, dan 135°. Adapun jarak yang digunakan tergantung dengan tingkat hubungan dari ketetanggaan antar piksel tersebut [4]. Metode GLCM dihitung menggunakan beberapa nilai yaitu:

- *Contrast*
Contrast merupakan sebuah variabel yang digunakan dalam GLCM sebagai pengukur intensitas kontras dengan piksel tetangga.
- *Correlation*
Correlation merupakan cara untuk mengukur tingkat korelasi piksel dengan nilai ketetanggaan.
- *Energi*
Energi merupakan jumlah elemen kuadrat pada GLCM yang sudah dinormalisasi.
- *Homogeneity*
Homogeneity merupakan nilai untuk mengukur kedekatan distribusi elemen dalam GLCM dengan diagonal GLCM.

2.4 K-Fold Cross Validation

Menurut K-Fold Cross Validation merupakan sebuah prosedur penyampelan ulang yang digunakan untuk mengevaluasi sebuah model *machine learning* pada sebuah data sampel yang terbatas. Prosedur ini memiliki sebuah parameter bernama *k* yang merepresentasikan jumlah pembagian dari kelompok data sampel [5]. Cara kerja *k-fold cross validation* adalah mengelompokkan data menjadi data latih dan data uji yang terpisah, yang kemudian dilakukan proses pengujian sebanyak jumlah parameter *k* kali.

2.5 Algoritma *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* adalah sebuah algoritma klasifikasi dengan algoritma klasifikasi yang dilakukan dari suatu data pembelajaran yang mempunyai jarak terdekat dengan objek tersebut [6]. Algoritma KNN merupakan algoritma klasifikasi yang menentukan kategori berdasarkan kategori terbanyak pada *K-Nearest Neighbor*. Dalam algoritma KNN terdapat sebuah nilai *K* sebagai suatu hal yang penting karena dapat mempengaruhi kerja dari algoritma *K-Nearest Neighbor* [7].

Prinsip kerja dari *K-Nearest Neighbor* adalah menghitung jauh atau dekatnya tetangga (dalam kasus ini yang mana berupa gambar, ialah piksel) yang dapat dihitung menggunakan jarak *Euclidean* dengan rumus:

$$D(a, b) = \sum_{k=1}^d (a_k - b_k)^2 \quad (1)$$

Keterangan:

$D(a, b)$ = Jarak.

d = Dimensi data.

k = Variabel data.

a_k = Data *testing*.

b_k = Data *sample*.

3 Metodologi Penelitian

3.1 Alur Penelitian

Dalam mencapai tujuan penelitian yang akan diteliti, penulis menyusun prosedur penelitian. Berikut merupakan kerangka berpikir dalam penelitian penulis:



Gambar 1. Alur Penelitian

3.2 Identifikasi Masalah

Penulis melakukan identifikasi masalah yang berkaitan dengan judul atau topik yang telah dipilih. Identifikasi yang akan digunakan selanjutnya menjadi dasar untuk memilih data maupun metode yang akan digunakan.

3.3 Studi Literatur

Pada tahap ini akan dilakukan proses perubahan besaran gambar pada citra asli untuk memudahkan input data citra pada program. Kemudian akan dilakukan perubahan data menjadi citra *grayscale*. Lalu dilakukan deteksi tepi *canny* untuk mendeteksi tepi objek pada citra. Kemudian dilakukan ekstraksi ciri citra menggunakan *Gray Level Co-occurrence Matrix* untuk mengambil data tekstur berupa *Contrast*, *Correlation*, *Energi*, dan *Homogenity*.

3.4 Akuisisi Data

Akuisisi data adalah tahapan untuk mengumpulkan data (dalam hal ini, citra) sebagai objek untuk penelitian. Pada penelitian ini, data yang akan diambil merupakan data citra daun tanaman zaitun dan daun tanaman mangga dan alpukat sebagai data pembandingan.

3.5 Pra Proses

Pada tahap ini akan dilakukan proses perubahan besaran gambar pada citra asli untuk memudahkan input data citra pada program. Kemudian akan dilakukan perubahan data menjadi citra *grayscale*. Lalu dilakukan deteksi tepi *canny* untuk mendeteksi tepi objek pada citra. Kemudian dilakukan ekstraksi ciri citra menggunakan *Gray Level Co-occurrence Matrix* untuk mengambil data tekstur berupa *Contrast*, *Correlation*, *Energi*, dan *Homogenity*.

3.6 Pembagian Data

Pada tahapan ini akan dilakukan pembagian data menjadi data latih dan data uji. Untuk pembagian data ini digunakan *K-Fold Cross Validation*. Cara ini membagi data dengan mencari parameter dari rata-rata nilai yang akan diteliti dan membaginya menjadi data yang dirasa cocok menjadi data uji dan data latih.

3.7 Klasifikasi

Pada tahap ini dilakukan klasifikasi data yang sudah dibagi menggunakan algoritma *K-Nearest Neighbor*. Data kan dikelompokkan menjadi 2 kelas yaitu kelas daun zaitun dan bukan daun zaitun. Algoritma akan mengklasifikasikan data citra uji ke dalam salah satu dari dua kelas tersebut.

3.8 Evaluasi

Pada tahap ini dilakukan evaluasi dengan melihat tingkat akurasi dari algoritma menggunakan metode *confusion matrix*. *Confusion matrix* digunakan untuk menghitung akurasi, presisi, dan *recall* dari hasil algoritma yang dijalankan. Berikut tabel dari *confusion matrix*.

		Nilai Sebenarnya	
		Positif	Negatif
Prediksi	Positif	TP	FP
	Negatif	FN	TN

Gambar 2. Tabel Confusion Matrix

Keterangan:

- TP = *True Positive* atau data kelas positif yang diprediksi benar.
- TN = *True Negative* atau data kelas negatif yang diprediksi benar.
- FP = *False Positive*, atau data kelas negatif yang salah prediksi.
- FN = *False Negative* atau data kelas positif yang salah prediksi.

Lalu dengan data di atas, didapatkan rumus untuk mencari akurasi, presisi, dan *recall* yang ditunjukkan oleh rumus dibawah.

1. Akurasi merupakan persentil rasio data yang diprediksi secara benar.

$$\frac{\Sigma \text{ Data Benar}}{n} \times 100\% \quad (2)$$

Keterangan:

- Σ Data Benar : Jumlah data benar yang diprediksi (TP+TN)
- n : Banyak data sampel yang ada

2. Presisi merupakan persentil rasio data benar positif terhadap kelas prediksi positif.

$$\frac{TP}{TP + FP} \times 100\% \quad (3)$$

3. *Recall* merupakan persentil rasio data benar positif terhadap kelas prediksi positif.

$$\frac{TP}{TP + FN} \times 100\% \quad (2)$$

Bila tingkat akurasi tidak sesuai dengan target yang sudah ditentukan akan dilakukan optimisasi algoritma agar dapat mencapai target.

3.9 Dokumentasi

Setelah semua tahapan berhasil dilaksanakan, dilakukan dokumentasi penelitian yang sudah selesai dalam bentuk sebuah laporan.

4 Hasil dan Pembahasan

Pada penelitian ini digunakan 50 daun zaitun dan 50 daun delima. Data citra diambil menggunakan kamera ponsel Redmi Note 10 dengan ukuran citra berupa 3000 x 4000 berformat JPEG (*.jpg). Dalam pemotretan dilakukan pengaturan EV atau exposure value yang lebih tinggi agar gambar terlihat lebih terang dan tidak ada *noise* pada *background* berwarna putih.



Gambar 4. Citra Daun Delima



Gambar 3. Citra Daun Zaitun

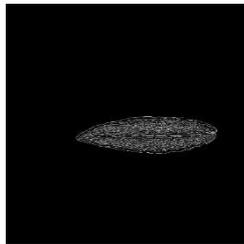
Kemudian dilakukan pengolahan pada citra yang akan digunakan untuk mempermudah proses ekstraksi ciri. Tahap pengolahannya terdapat tiga, yaitu pengubahan ukuran citra, segmentasi citra, pengubahan citra menjadi *grayscale*, dan deteksi tepi *canny*. Lalu citra asli akan diubah skalanya menjadi 1368 x 912 piksel untuk memudahkan input gambar ke dalam program. Kemudian citra akan dipotong sebesar 912 x 912 piksel dengan koordinat potong yaitu $x = 228$ piksel dan $y = 0$ piksel untuk memperkecil *background* gambar yang tidak dipakai.

Tahap selanjutnya dilakukan proses *grayscale* yang menjadi sebuah syarat mutlak untuk melakukan ekstraksi ciri GLCM. Adapun langkah yang dilakukan ialah mengubah intensitas cahaya yang ada pada citra menjadi hitam putih pada *matlab*.



Gambar 5. Citra hasil grayscale

Langkah akhir yaitu dilakukan deteksi tepi pada citra menggunakan metode *canny*. Hal ini dilakukan supaya objek citra yang diteliti dapat terlihat detail dan bentuk dari setiap tekstur yang ada.



Gambar 6. Citra hasil tepi Canny

Kemudian dilakukan ekstraksi ciri tekstur menggunakan metode *Gray Level Co-occurrence Matrix* (GLCM). Langkah yang harus dilakukan pada tahap ini adalah mengambil fitur tekstur pada citra dengan nilai jarak piksel (*pixel distance*) sebesar 1 yang lalu akan diambil 4 variabel data yaitu *contrast*, *correlation*, *energi*, dan *homogeneity*. Beberapa contoh nilai rata-rata dari hasil ekstraksi ciri GLCM akan ditampilkan pada tabel berikut:

Tabel 1. Contoh data hasil GLCM

Objek	File	Contrast	Correlation	Energy	Homogeneity
Daun Delima	img_01	0.009545	0.429938	0.973804	0.995228
	img_02	0.016079	0.328013	0.960252	0.99196
	img_03	0.012271	0.422783	0.966622	0.993865
	img_04	0.014033	0.386378	0.963296	0.992984
Daun Zaitun	img_51	0.007326	0.494569	0.978232	0.996337
	img_52	0.007365	0.52829	0.977075	0.996317
	img_53	0.006611	0.473763	0.980869	0.996694
	img_54	0.012093	0.428218	0.966903	0.993953

Data yang telah diekstrak cirinya harus dinormalisasi terlebih dahulu untuk mempermudah proses saat melakukan klasifikasi. Data yang di normalisasi dibagi menjadi dua kelompok yaitu data latih dan data uji. Metode dalam pembagian data ini menggunakan *K-Fold Cross Validation*. Metode ini digunakan untuk menemukan parameter yang optimal dalam melakukan klasifikasi. Penerapan metode ini akan dilakukan dengan membagi data tiga kali dengan memberi input nilai konstanta Kf dengan nilai 4, 5, dan 10.

Setelah pembagian data dilakukan, data yang sudah dibagi akan diberi label data *training* yang akan digunakan sebagai data untuk melatih algoritma *K-Nearest Neighbor* (KNN). Algoritma *K-Nearest Neighbor* merupakan suatu metode klasifikasi yang dilakukan dari suatu data pembelajaran yang mempunyai jarak terdekat dengan objek tersebut. Jadi data *testing* atau uji akan diklasifikasikan berdasarkan pada model yang sudah dijalankan data *training* atau latih yang memiliki ketetanggaan terdekat dengan data uji dan masuk ke dalam kategori atau kelas yang paling banyak ada dan cocok dengan nilai ketetanggaan terdekatnya (K). Selanjutnya, dilakukan evaluasi data menggunakan perhitungan akurasi, presisi, dan *recall* menggunakan *confusion matrix*. Berikut adalah perbandingan dari seluruh hasil evaluasi klasifikasi yang telah dilakukan:

Tabel 2. Hasil evaluasi klasifikasi

K-Fold	KNN	Akurasi (%)	Presisi (%)	Recall (%)
4	3	87	93.9	85.9
4	5	89	88.3	91.05
4	7	87.85	90.05	88.55
5	3	88	86	90.12
5	5	87	86	87.62
5	7	85	85	89.6
10	3	86	79	88.33
10	5	85	74	86.83
10	7	85	67	88.5
Rata-rata		77.985	74.925	79.65

Pada tabel di atas menunjukkan seluruh nilai akurasi, presisi, dan *recall* dari masing-masing parameter. Nilai K pada *K-Fold* yang dipasangkan dengan nilai KNN berjumlah masing-masing 4 dan 5 memiliki akurasi terbesar dengan nilai 89%. Namun, dari kesembilan parameter yang mungkin, selisih di antaranya tidak terlalu besar walau pada nilai K untuk *K-Fold* sama dengan 10 didapati nilai presisi yang lebih rendah ketimbang parameter kemungkinan lainnya.

5 Simpulan

Berdasarkan penelitian yang sudah dilakukan, dapat ditarik kesimpulan bahwa metode ekstraksi ciri *Gray Level Co-occurrence Matrix* (GLCM) dapat digunakan untuk mengambil fitur ciri tekstur pada objek citra berupa citra daun zaitun dengan baik menggunakan parameter nilai *contrast*, *correlation*, *energi*, dan *homogeneity* yang dapat membantu klasifikasi menggunakan algoritma KNN dengan rata-rata akurasi di atas 80%. Kemudian performa Performa dari algoritma *K-Nearest Neighbor* untuk pengklasifikasian tanaman zaitun dan delima dapat dikatakan baik dengan nilai evaluasi pada akurasi dengan nilai 89%, presisi sebesar 88.3%, dan *recall* sebesar 91.05% dengan skala perbandingan data 75:25 menggunakan *4-fold Cross Validation* dan nilai konstanta K pada KNN sebesar 7. Dan nilai K dan jumlah *K-fold* yang semakin tinggi tidak selalu berbuah dengan nilai akurasi yang tinggi, malah cenderung merusak kinerja dari algoritma KNN seperti yang dapat dilihat pada *10-fold Cross Validation* yang akurasi dan presisi terendah di antara parameter lainnya.

Terdapat pula saran yang dapat yang dapat digunakan untuk pengembangan pada penelitian ini ke depannya, seperti Menambahkan data tanaman lainnya selain tanaman delima sebagai faktor klasifikasi yang lebih baik, menambahkan data tanaman zaitun lainnya seperti objek buah atau batang dari tanaman, menambahkan fitur ekstraksi ciri lainnya seperti warna objek ataupun yang lainnya, dan mengembangkan penelitian dengan membuat sistem kecerdasan buatan sehingga sistem dapat mempelajari sendiri agar sistem dapat mengklasifikasikan dengan lebih pintar.

Referensi

- [1] R. Govaerts, P.S.Green (2021). Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; <http://wesp.science.kew.org/> Retrieved 2 July 2021.
- [2] Green', P. S. (2002) 'Type species: *Olea europaea* L. (lectotype), Green & Wickens in Kit Tan', Kew Bulletin, 57(1), p. 8.
- [3] Hastuti, I. (2016) "Perbandingan Metode Deteksi Tepi Menggunakan Metode Canny, Prewitt Dan Sobel Pada Image Ikan", *Jurnal Repositori Poliban*, 1(Vol 1 (2016): Inovasi Teknologi Tepat Guna untuk Mewujudkan Sinergi Perguruan Tinggi dengan Masyarakat), pp. A129–A137.
- [4] Arrozi, A.Z., (2019). *Klasifikasi Daun Menggunakan Metode Grey Level Co-Occurance Matrix (GlcM) Dan Algoritma Learning Vector Quantization (LVQ)* Doctoral dissertation, Universitas Pembangunan Nasional Veteran Jakarta.
- [5] Browniee, Jason, 2018. *A Gentle Introduction to k-Fold Cross Validation, Statistical Methods*. Retrieved from <https://machinelearningmastery.com/k-fold-cross-validation/> on 1 April 2021.
- [6] Andono, P. N., Sutojo, T., & Mulijono. (2017) "Pengelolaan Citra Digital", Yogyakarta: Penerbit ANDI.
- [7] Banjarsari, M. A., Budiman, H. I., & Farmadi, A. (2015) "Penerapan K-Optimal Pada Algoritma KNN untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer FMIPA UNLAM Berdasarkan IP Sampai Dengan Semester 4" *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 2(2), 50—64. ISSN: 2406-7857.